iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** III **Month of publication:** March 2022

DOI: https://doi.org/10.22214/ijraset.2022.41065

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com

Review Paper on Detection of Malicious URLs Using Machine Learning Techniques

Miss. Mayuri Arvind Pohane[1], Dr. A. A. Bardekar[2]
[1]*PG Scholar,* [2]*Professor, Computer Science & Engineering, Sipna College Of Engineering And Technology, Amravati, Maharashtra, India*

Abstract: ***Malicious websites are most serious threats over the Web. Ever since the inception of the internet, there has been a rise in malicious content over the web such has terrorism, financial fraud, phishing and hacking that targets user's personal information.***
Till today the various systems have been invent for the detection of a malicious website based on keywords and data content of the websites.
This existing method have some drawbacks results into numbers of victims to increase. Hence we developed a system which helps the user to identify whether the website is malicious or not. Our system identifies whether the site is malicious or not through URL.
The proposed system is fast and more accurate compared to current system. The classifier is trained with datasets of 1000 malicious sites and 1000 legitimate site URLs. Trained classifier is used for detection of malicious URLs.
Keywords: Malicious URLs, Classifier, Feature Extraction, ID3 Algorithm

I. INTRODUCTION

Since there is rapid growth to surf over internet, there has been a rise in malicious content over the web which leak user's confidentiality or circulating unauthorized data. Similarly, the quality and quantity of web attacks have increased. There is no doubt that harmful websites can be extremely damaging all organizations and access user's information in order to gain access to that business network.
The number of attacks has increased three to five-fold in less than 5 years, resulted loss is billions and the number of the malicious websites is raising day by day.
The current system which is used for detection of malicious websites is not effective in detecting the temporary or new malicious websites. In this paper, we propose a system which uses an automated classifier for the detection of malicious websites using URL features

II. LITERATURE SURVEY

1) They will conduct an experimental user study to evaluate the effectiveness of image comparison and display customization as techniques for users to identify remote servers. We are currently designing a between-subjects study to compare our prototype to other techniques. In the study, participants will be asked to create an account on a remote server and to login. We will periodically send the users email that asks them to login to the website in order manage their funds (which is their payment for participation in the study). Occasionally, users will be sent to a website that spoofs the content of the site as well as the security indicators (such as their trusted window). Participants will be divided into three groups, one using Dynamic Security Skins, one using a shared secret scheme and another using only a standard SSL equipped browser (the control condition). Effectiveness of the prototype will be measured by the performance and error rate in account creation and login tasks, the ability for users to authenticate legitimate servers, the rate of detecting spoof attempts and user satisfaction. Additionally, we plan to release the application to the public for widespread testing
2) This survey provides a system review of extensive research on phishing techniques and countermeasures. Previous surveys and taxonomies either concentrate on one specific aspect of phishing such as anti-phishing tools (Abbasi et al. 2010; Zhang et al. 2011a), or fail to provide an integrated overview of research approaches to various phishing techniques (Huajun et al. 2009; Wetzel 2005; Ollmann 2007a); The taxonomy proposed in this research is multi-dimensional, which distinguishes itself from the previous ones that are focused on a single dimension. In addition, the phishing environment covered in existing taxonomies is limited to traditional channels such as e-mails and spoofed websites

3) They have described an approach for classifying URLs automatically as either malicious or benign based on supervised learning across both lexical and host-based features. We argue that this approach is complementary to both blacklisting — which cannot predict the status of previously unseen URLs — and systems based on evaluating site content and behavior — which require visiting potentially dangerous sites. Further, we show that with appropriate classifiers it is feasible to automatically sift through comprehensive feature sets (i.e., without requiring domain expertise) and identify the most predictive features for classification. An open issue is how to scale our approach to handle millions of URLs whose features evolve over time. We address the issue in subsequent work by using online learning algorithms.

4) They have developed a system for detection of malicious websites through URL which based on an automated classifier. The classifier is trained with the dataset of legitimate and malicious websites. The trained classifier is for the detection of any URL. Further, the accuracy of thesystem increases as the classifier is trained with more data set.

III. SYSTEM DIAGRAMS

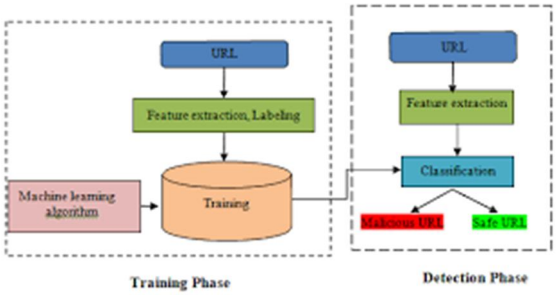


Fig : Learning to detect malicious web sites from suspicious URLs.

IV. CONCLUSION

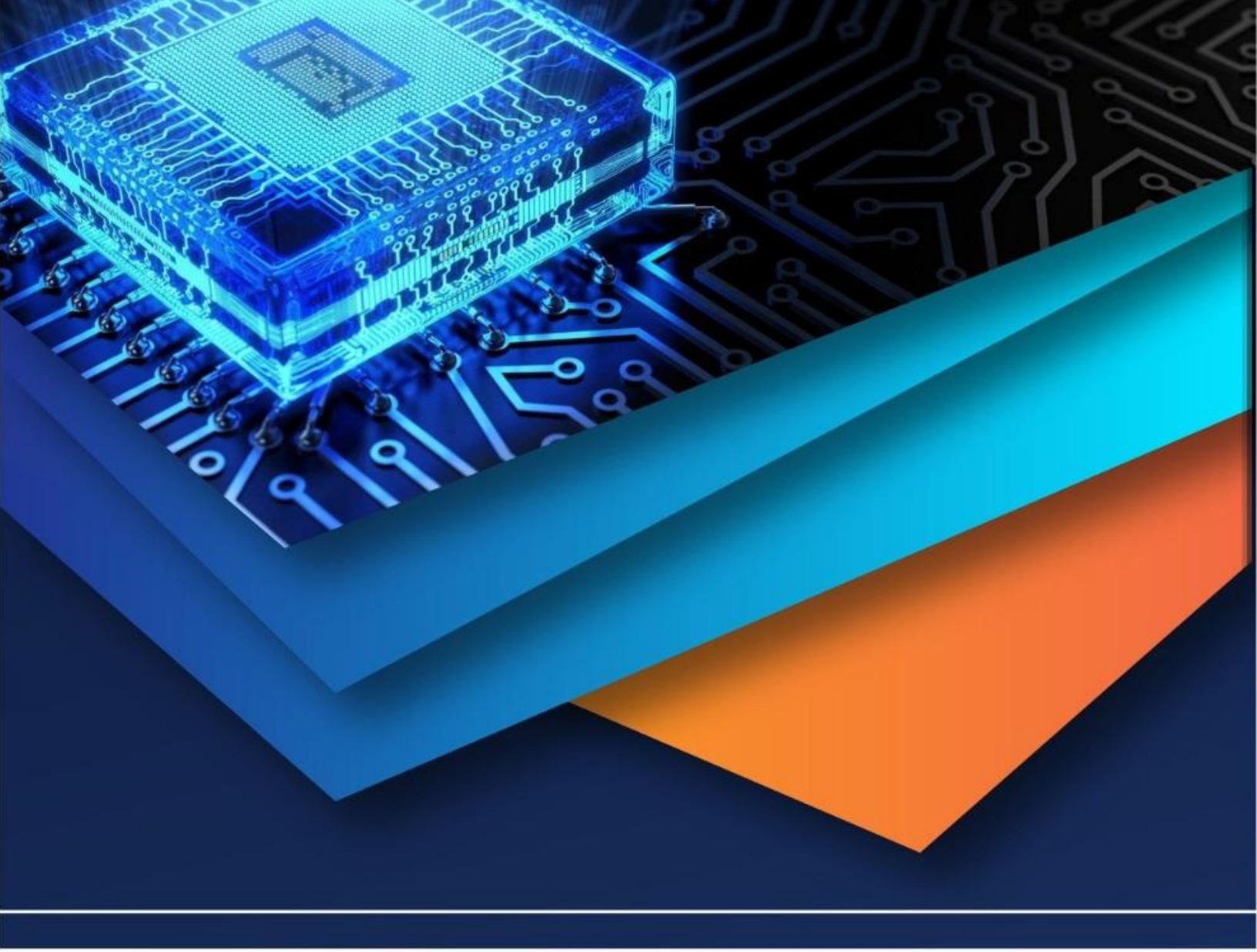
Thus we studied the above literature survey on that conclusion we design a system for detection of malicious websites through URL which based on an automated classifier. The classifier is trained with the dataset of legitimate and malicious websites. The trained classifier is for the detection of any URL. Further, the accuracy of the system increases as the classifier is trained with more data set.

V. ACKNOWLEDGEMENT

First and foremost, I would like to express my sincere gratitude to my **PROF. A.A. Bardekar** who has in the literal sense, guided and supervised me. I am indebted with a deep sense of gratitude for the sconstant inspiration and valuable guidance throughout the work.

REFERENCES

[1] S. Dhamija, R., and Tygar, J., “The battle against phishing: Dynamic security skins”, In Proc. ACM Symposium on Usable Security and Privacy (SOUPS 2005), pp.77–88,.
[2] Chandrasekaran, M., Narayanan, K., and Upadhyaya, S., “Phishing email detection based on structural properties”, Proceedings of the NYS Cyber Security Conference, 2006.
[3] J. Ma, L.K. Saul, S. Savage, G.M. Voelker, “Beyond blacklists: Learning to detect malicious web sites from suspicious URLs”, In: Proc. 15th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, Paris, France, 2009, pp. 1245-1254.
[4] Yogesh Dubey , Palghar Pranil Chaudhari Student ,Tina D’abreo Lecturer Efficient Detection of Legitimate and Malicious URLs using ID3 Algorithm
[5] Ma, Justin, et al. "Beyond Blacklists: learning to detect malicious websites from suspicious URLs." Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2009
[6] Jin-Lee Lee,Doung-Hyun Kim,Chang-Hoon Lee. “Heuristic-based Approach for Phishing Site Detection Using URL Features” Third Intl. Conf. on Advances in Computing, Electronics and Electrical Technology - CEET 2015.
[7] Sana Ansari and Jayant Gadge. “Architecture for Checking Trustworthiness of Websites “International journal of computer application, Volume 44, April 2012
[8] Mustafa Aydin and Nazife Baykal “Feature Extraction and Classification Phishing Websites Based on URL” Cyber Defence and Security Laboratory of METUCOMODO, IEEE CNS 2015.
[9] Nguyen, Luong Anh Tuan, et al. "A novel approach for phishing detection using URL-based heuristic." Computing, Management and Telecommunications (ComManTel), 2014 International Conference on. IEEE, 2014.
[10] Canali, Davide, et al. "Prophiler: a fast filter for the large-scale detection of malicious web pages." Proceedings of the 20th international conference on World wide web. ACM, 2011.
[11] Sumalatha Ramachandran, Sujaya Paulraj, Sharon Joseph and Vetriselvi Ramaraj, “Enhanced Trustworthy and High-Quality Information Retrieval System for Web Search Engines”, IJCSI International Journal of Computer Science Issues, Vol. 5, October 2009, pp38- 42.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY